

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
30 November 2000 (30.11.2000)

PCT

(10) International Publication Number
WO 00/72449 A1

(51) International Patent Classification⁷: H03M 13/23, 13/37

(74) Agent: PATENTTITOIMISTO TEKNOLOGIS KOLSTER OY; c/o Kolster OY AB, Iso Roobertinkatu 23, P.O. Box 148, FIN-00121 Helsinki (FI).

(21) International Application Number: PCT/FI00/00457

(22) International Filing Date: 19 May 2000 (19.05.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
991159 21 May 1999 (21.05.1999) FI

(71) Applicant (for all designated States except US): NOKIA MOBILE PHONES LTD [FI/FI]; Keilalahdentie 4, FIN-02150 Espoo (FI).

(72) Inventors; and

(75) Inventors/Applicants (for US only): HÄMÄLÄINEN, Ari [FI/FI]; Linnoituskuja 1 C 19, FIN-01280 Vantaa (FI). HENRIKSSON, Jukka [FI/FI]; Kurkijoentie 19 C, FIN-02140 Espoo (FI).

(81) Designated States (national): AE, AG, AL, AM, AT, AT (utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, CZ (utility model), DE, DE (utility model), DK, DK (utility model), DM, DZ, EE, EE (utility model), ES, FI, FI (utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KR (utility model), KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

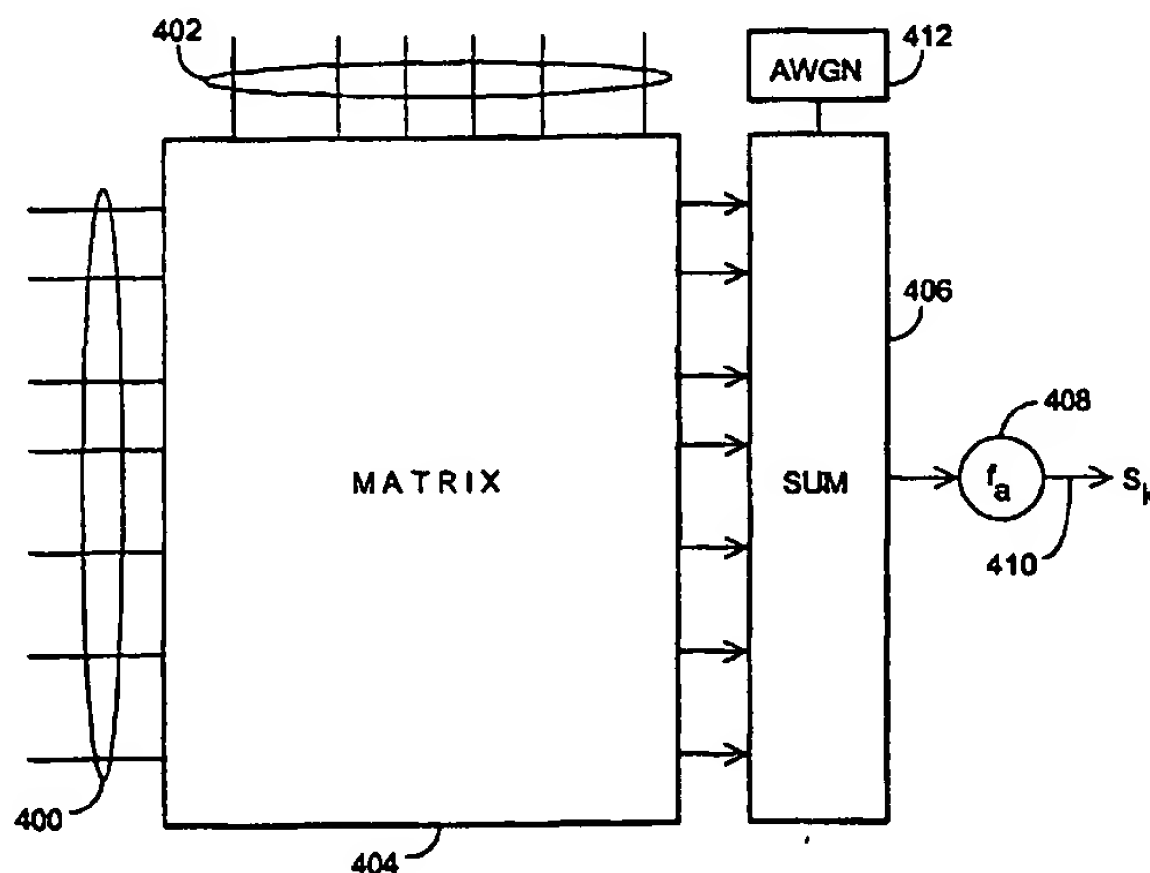
(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— With international search report.

[Continued on next page]

(54) Title: METHOD AND ARRANGEMENT FOR IMPLEMENTING CONVOLUTIONAL DECODING



(57) Abstract: The invention relates to a method and arrangement for advantageously decoding and channel correcting a convolutionally encoded signal received over a transmission path. The signal comprises code words and the arrangement comprises a neural network comprising a set of neurons which comprise a set of inputs and an output. The received code words (400) and at least some of the output signals (402) of the neural network neurons are connected to the inputs of the neurons, and the neurons comprise means (404) for multiplying at least some of the neuron inputs prior to combining means (406). The arrangement also comprises means (123) for estimating the transmission channel. Further, estimated channel data (400) is connected to the inputs of the neurons and a predetermined neuron is arranged to give an estimate of the channel-corrected and decoded symbol in its output signal.

WO 00/72449 A1



— Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHOD AND ARRANGEMENT FOR IMPLEMENTING CONVOLUTIONAL DECODING

FIELD OF THE INVENTION

The invention relates to a method and an arrangement for decoding
5 and channel correcting a convolutionally encoded signal in a receiver, by means of a feedback neural network, said signal being received over a transmission path.

BACKGROUND OF THE INVENTION

In telecommunication connections, the transmission path used for
10 transmitting signals is known to cause interference to telecommunication. This occurs regardless of the physical form of the transmission path, i.e. whether the transmission path is a radio link, an optical fibre or a copper cable. Particularly in telecommunications over a radio path there are frequently situations where the quality of the transmission path varies from one occasion to another
15 and also during a connection.

Radio path fading is a typical phenomenon that causes changes in a transmission channel. Other simultaneous connections may also cause interferences and they can vary as a function of time and place.

In a typical radio telecommunication environment the signals between a transmitter and a receiver propagate on a plurality of routes. This
20 multipath propagation mainly results from signal reflections from surrounding surfaces. The signals that propagated via different routes reach the receiver at different times due to different transit delays. Various methods have been developed in order to compensate the interference caused by this multipath
25 propagation.

In order to reduce the effects of interference caused by the transmission path, a digital signal is encoded so as to make the transmission connection more reliable. Thus the errors caused by the interference in the transmitted signal can be detected and, depending on the encoding method used,
30 also corrected without retransmission. Conventional encoding methods used in digital telecommunication include, for instance, block coding and convolutional coding. In block coding the bits to be encoded are grouped into blocks at the end of which are added parity bits, by means of which the correctness of the bits of the preceding block can be checked.

35 In the receiver, the errors caused by multipath propagation are typi-

cally first corrected, for instance, by a linear transversal corrector and thereafter the convolutional code is decoded.

The efficiency of convolutional coding depends on the code rate and constraint length used. A large constraint length enables good error correction capability, but on the other hand, decoding by known methods is then
5 very complicated.

In general, convolutional coding is decoded by using a Viterbi algorithm that has nearly optimal performance. However, the Viterbi algorithm has a drawback that its complexity increases exponentially as the constraint length
10 increases. This restricts the constraint lengths available.

Another known decoding method is sequential decoding that is described in greater detail in *Digital Communications*, J.G. Proakis, 3rd edition, pp. 500-503. One drawback of this algorithm is that the decoding delay does not remain constant but varies.

15 Yet another known decoding method is a so-called stack algorithm that is described in greater detail in the above-mentioned publication *Digital Communications*, J.G. Proakis, 3rd edition, pp. 503-506. The performance of this algorithm is not so good as that of the Viterbi algorithm.

Of the known methods, the Viterbi algorithm has the best performance for decoding the convolutional code, but its implementation has turned
20 out to be extremely difficult in practice as the constraint length increases. The difficult implementation of the complicated Viterbi algorithm by circuitry has restricted the constraint lengths available for use.

A separate channel corrector and decoder are a suboptimal solution. Use of channel data, in particular within the Viterbi algorithm when computing the metrics, leads to increased complexity and the implementation of practical applications is impossible.

BRIEF DESCRIPTION OF THE INVENTION

The object of the invention is to provide a method and an arrangement
30 implementing the method such that the above-mentioned drawbacks can be solved. This is achieved by a method of the invention for decoding a convolutionally encoded signal received over a transmission path, which signal comprises code words and in which method a transmission channel is estimated, decoding is carried out by means of an artificial neural network, the
35 neural network comprising a set of neurons which comprise a set of inputs and an output, the received code word set is conducted to the inputs of the neuron,

at least some of the inputs of the neuron are multiplied, after multiplication some of the inputs of the neuron are combined, some of the output signals of the neural network neurons are fed back to the inputs of each neuron, initial values of the neurons are set and the network is allowed to stabilize.

5 Further, in the method of the invention the multiplication of the neuron inputs depends on the convolutional code used in signal encoding and on the estimated channel and on the fact that an estimate of the decoded and channel-corrected symbol is conducted from the output signal of a predetermined neuron to the output of the network after the network has reached a
10 stabilized state, the set of code words in the shift register is updated and the above-described four last steps are repeated until the desired number of code words are decoded.

The invention also relates to an arrangement for decoding and channel correcting a convolutionally encoded signal received over a trans-
15 mission path, the signal comprising code words and the arrangement comprising a neural network which comprises a set of neurons which comprise a set of inputs and an output, the received code words being applied to the inputs of said neurons and to at least some of the output signals of the neural network neurons, the neurons comprising means for multiplying at least some of the
20 inputs of the neuron prior to combining means, the arrangement comprising means for estimating the transmission channel. Further in the arrangement of the invention, estimated channel data is applied to the inputs of the neurons, and a predetermined neuron is arranged to give an estimate of the channel-corrected and decoded symbol in its output signal.

25 The preferred embodiments of the invention are disclosed in the dependent claims.

In the solution of the invention, convolutional code decoding and channel correction are performed by means of a feedback neural network. Several advantages are achieved with the solution of the invention. Perform-
30 ance that is close to that of the Viterbi algorithm is achieved with the solution of the invention by means of considerably simpler circuitry. In the solution of the invention, equally complicated circuitry permits a larger constraint length and thus improved error correction over the Viterbi algorithm.

The neural network of the invention can be readily constructed by
35 semiconductor technology, since the neurons of the neural network are very similar to one another in structure, only the input couplings vary. Conse-

quently, to design and implement even a large network does not involve a great amount of work. The solution can also be advantageously implemented as a VLSI circuit, which makes it fast.

BRIEF DESCRIPTION OF THE DRAWINGS

5 In the following the invention will be described in greater detail in connection with preferred embodiments with reference to the accompanying drawings, wherein

Figure 1 illustrates an example of a telecommunications system where the solution of the invention can be applied,

10 Figure 2 shows an example of a neuron designed by means of the method of the invention,

Figure 3 shows an example of couplings of the neural network neuron designed by means of the method of the invention,

15 Figure 4 illustrates an example of the implementation of the neural network neuron of the invention,

Figures 5a and 5b illustrate an implementation example of a neural network of the invention,

Figure 6 illustrates an example where decoding is performed by means of a plurality of parallel neural networks.

20 DETAILED DESCRIPTION OF THE INVENTION

The solution of the invention can be applied in any digital data transmission system where a signal to be transmitted is encoded by convolutional coding. Examples of possible systems include cellular radio systems, radio links, wired networks or optical networks.

25 Figure 1 illustrates an example of a telecommunications system according to the invention. The figure shows a transmitter 100, a signal transmission path 102 and a receiver 104. The operations of the invention are implemented in the receiver. The signal transmission path can be a radio path, a copper cable or an optical fibre, for instance.

30 In the transmitter, a data source 106 generates a signal 108 composed of bits. An error correcting code can be added to the data source signal 108 by a known FEC (Forward Error Correcting) method, if necessary. This is not shown in the figure. The signal 108 is applied to an encoder 110 where convolutional coding is effected with desired parameters (constraint length and
35 code rate). The encoder can have a prior art structure. The encoded signal

can be understood as code words whose length depends on the code used. The encoded signal is applied to a modulator 112 where the desired modulation is performed. The modulated signal is further applied to output stages 114 and a channel adapter 116 of the transmitter which are system-dependent components. The output stage 114 can be a radio frequency amplifier and a filter, for instance. The channel adapter 116 can be an antenna or an electrical/opto-converter, for instance.

A transmitter output signal 118 propagates through the transmission path 102, i.e. the channel, to the receiver 104. The receiver comprises a channel adapter 120 and a receiver pre-stage 122 which are system-dependent components. The output stage 122 can comprise a radio frequency amplifier and a filter, for instance. The channel adapter 120 can comprise an antenna or an opto/electrical converter, for instance. The signal is applied from the pre-stage 122 further to a channel estimator 123 where channel estimation is performed by using known methods. From the channel estimator, the signal is applied to a demodulator 124 from which is typically received a baseband bit stream 126 which can be complex and comprise separate I- and Q-branch bits. Timing information is derived from the received signal in a regenerator 128 by a known method. This timing information is used in demodulation. The receiver further comprises a control unit 130 which can be advantageously implemented by a processor. The control unit 130 controls the operation of the receiver. Information on the timing is also applied to the control unit.

The signal 126 coming from the demodulator 124 thus comprises the demodulated bits which can be presented as code words. The code words are applied to a decoder 132 which is implemented by a neural network in the receiver of the invention. Convolutional coding is decoded in the decoder and the decoder output signal 134 corresponds to the original signal 108. In connection with convolutional coding, the channel data obtained from the channel estimator 123 is taken into account, i.e. channel correction is performed. If an error correcting code was used in the transmitter, it is decoded at this stage. This is not shown in the figure.

The method for decoding the convolutional code and performing the channel correction in accordance with the invention is studied next in greater detail. It is assumed in this example that 4 QAM modulation is used. Generally, the aim of the decoding is to find a bit sequence B that minimizes the function

$$\min_B \sum_{s=0}^T \| r(t + sdt) - h\gamma_c(B(t + sdt)) \|^2 =: \min_B f(B) \quad (1)$$

where $r(t + sdt)$ is a received code word at a time instant $t + sdt$,
 dt is a symbol interval,
 5 $h = [h_1, h_2]$ are channel coefficients of a two-path channel,
 $\gamma_c(B(t + sdt)) = [\gamma(B(t + sdt)), \gamma(B(t + (s-1)dt))]$ a vector with a current
 code word and a preceding code word as components. When 4
 QAM modulation is used, these are complex values.

10 For the sake of clarity, $r(s) = r(t + sdt)$ and $B(s) = B(t + sdt)$ are used.
 Furthermore, instead of bit values 0 and 1, values -1 and 1 are used. In For-
 mula (1)

$$B(s) = [b(s-L+1), \dots, b(s)], \quad b \in \{-1, 1\},$$

$b(-L+1), \dots, b(0)$ are fixed, L is the constraint length and T is the decoding de-
 15 lay. The encoder function γ corresponds to the generation matrix G of a con-
 volutional encoder. On the transmission path, interference, such as noise, ap-
 pears in the signal. A noisy code word received at a time instant s is indicated
 by $r(s)$. It is assumed that the noise is white Gaussian noise.

Let us study here the generation of a neural network of the inven-
 20 tion by means of an example employing a convolutional code with code rate $\frac{1}{2}$
 and constraint length $L = 3$. The generation matrix G for such a code is

$$G = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

The encoder for the convolutional code with code rate $\frac{1}{2}$ and con-
 25 straint length L can be written as a complex function

$$\gamma(B(s)) = \gamma_1(B(s)) + i \gamma_2(B(s)),$$

where

$$\gamma_j(B(s)) = \prod_{i=1}^L b(s+1-i)^{g_{ji}} (-1)^{g_{ji}},$$

and g_{ji} s are the elements of the matrix G . Now, as in this example $L = 3$, the
 30 encoder is of the form

$$\gamma(B(s)) = -b(s)b(s-2) + i b(s)b(s-1)b(s-2)$$

and the received noisy complex code word is $r(s) = r_1(s) + i r_2(s)$.

Equation (1) is written out as:

$$\sum_s \| r(s) - h\gamma_c(B(s)) \|^2 = \quad (2)$$

$$\begin{aligned}
& \sum_s \left(\frac{1}{2} [re(r(s))^2 + im(r(s))^2] \right. \\
& \quad + [re(r(s)h_1^c)b(s)b(s-2) - im(r(s)h_1^c)b(s)b(s-1)b(s-2)] \\
& \quad + [re(r(s)h_2^c)b(s-1)b(s-3) - im(r(s)h_2^c)b(s-1)b(s-2)b(s-3)] \\
& \quad + \frac{1}{2} h_1^c h_1 [b(s)^2 b(s-2)^2 + b(s)^2 b(s-1)^2 b(s-2)^2] \\
& \quad + \frac{1}{2} h_2^c h_2 [b(s-1)^2 b(s-3)^2 + b(s-1)^2 b(s-2)^2 b(s-3)^2] \\
& \quad + re(h_1^c h_2)b(s)b(s-1)b(s-2)b(s-3) - im(h_1^c h_2)b(s)b(s-1)b(s-2)b(s-3) \\
& \quad \left. + im(h_1^c h_2)b(s)b(s-1)b(s-2)^2 b(s-3) + re(h_1^c h_2)b(s)b(s-1)^2 b(s-2)^2 b(s-3) \right).
\end{aligned}$$

In the above, the superscript c refers to a complex conjugate. As $b(s-1)^2 = 1$ and $b(s-2)^2 = 1$, the minima of the above function (2) and the function

$$\begin{aligned}
5 \quad \hat{f}(B) := & \sum_s \left(\frac{1}{2} [re(r(s))^2 + im(r(s))^2] \right. \\
& \quad + [re(r(s)h_1^c)b(s)b(s-2) - im(r(s)h_1^c)b(s)b(s-1)b(s-2)] \\
& \quad + [re(r(s)h_2^c)b(s-1)b(s-3) - im(r(s)h_2^c)b(s-1)b(s-2)b(s-3)] \\
& \quad + \frac{1}{2} h_1^c h_1 [b(s)^2 b(s-2)^2 + b(s)^2 b(s-1)^2 b(s-2)^2] \\
& \quad + \frac{1}{2} h_2^c h_2 [b(s-1)^2 b(s-3)^2 + b(s-1)^2 b(s-2)^2 b(s-3)^2] \\
& \quad + re(h_1^c h_2)b(s)b(s-1)b(s-2)b(s-3) - im(h_1^c h_2)b(s)b(s-1)b(s-2)b(s-3) \\
& \quad \left. + im(h_1^c h_2)b(s)b(s-1)b(s-3) + re(h_1^c h_2)b(s)b(s-3) \right). \tag{3}
\end{aligned}$$

are the same. The differences between the functions are in the two last terms.

To eliminate complicated feedbacks, the function (3) will be employed.

(3) is differentiated with respect to $b(k)$. Then we obtain a partial derivative which is of the form

$$\frac{\partial \hat{f}(B)}{\partial b(k)} = \tag{4}$$

$$\begin{aligned}
& re(r(k)h_1^c)b(k-2) - im(r(k)h_1^c)b(k-1)b(k-2) \\
& - im(r(k+1)h_1^c)b(k+1)b(k-1) + re(r(k+1)h_2^c)b(k-2) \\
& - im(r(k+1)h_2^c)b(k-1)b(k-2) + re(r(k+2)h_1^c)b(k+2) \\
& - im(r(k+2)h_1^c)b(k+2)b(k+1) - im(r(k+2)h_2^c)b(k+1)b(k-1) \\
& + re(r(k+3)h_2^c)b(k+2) - im(r(k+3)h_2^c)b(k+2)b(k+1) + re(h_1^c h_2)b(k-3) \\
& + re(h_1^c h_2)b(k-2)b(k-1)b(k-3) - im(h_1^c h_2)b(k-2)b(k-3) \\
& + im(h_1^c h_2)b(k-1)b(k-3) + re(h_1^c h_2)b(k+1)b(k-1)b(k-2) \\
& + im(h_1^c h_2)b(k+1)b(k-2) + re(h_1^c h_2)b(k+2)b(k+1)b(k-1) \\
& - im(h_1^c h_2)b(k+2)b(k-1) + re(h_1^c h_2)b(k+3)b(k+2)b(k+1) \\
& - im(h_1^c h_2)b(k+3)b(k+1) + im(h_1^c h_2)b(k+3)b(k+2) + re(h_1^c h_2)b(k+3) \\
& + 5(h_1^c h_1 + h_2^c h_2)b(k).
\end{aligned}$$

Now we use a gradient method

$$b(k) = b(k) - \alpha \frac{\partial \hat{f}(B)}{\partial b(k)}.$$

5 By selecting

$$\alpha = \frac{1}{5(h_1^c h_1 + h_2^c h_2)}$$

it is possible to eliminate the feedbacks appearing in the last term of Formula (4). Further by assuming that $b(k)$ obtains values -1 or 1, the following equivalence is true for all the values $k = 0, \dots, T$

$$\begin{aligned}
b(k) = \text{sign}(& - re(r(k)h_1^c)b(k-2) + im(r(k)h_1^c)b(k-1)b(k-2) \\
& + im(r(k+1)h_1^c)b(k+1)b(k-1) - re(r(k+1)h_2^c)b(k-2) \\
& + im(r(k+1)h_2^c)b(k-1)b(k-2) - re(r(k+2)h_1^c)b(k+2) \\
& + im(r(k+2)h_1^c)b(k+2)b(k+1) + im(r(k+2)h_2^c)b(k+1)b(k-1) \\
& - re(r(k+3)h_2^c)b(k+2) + im(r(k+3)h_2^c)b(k+2)b(k+1) - re(h_1^c h_2)b(k-3) \quad (5) \\
& - re(h_1^c h_2)b(k-2)b(k-1)b(k-3) + im(h_1^c h_2)b(k-2)b(k-3) \\
& - im(h_1^c h_2)b(k-1)b(k-3) - re(h_1^c h_2)b(k+1)b(k-1)b(k-2) \\
& - im(h_1^c h_2)b(k+1)b(k-2) - re(h_1^c h_2)b(k+2)b(k+1)b(k-1) \\
& + im(h_1^c h_2)b(k+2)b(k-1) - re(h_1^c h_2)b(k+3)b(k+2)b(k+1) \\
& + im(h_1^c h_2)b(k+3)b(k+1) - im(h_1^c h_2)b(k+3)b(k+2) - re(h_1^c h_2)b(k+3) \quad).
\end{aligned}$$

From the above the following updating rule is obtained: one k at a time is selected and a corresponding bit is updated by using Formula (5) and this is repeated until no changes occur in the bits, i.e. the network is stabilized. It should be noted that if $k = T-1$, or $k = T$, then $b(T+1) = 0$ and $b(T+2) = 0$ are set. In the actual implementation, the operation may take place in parallel such that all neurons k are updated simultaneously.

The iteration according to Formula (5) can thus be presented as an artificial neuron of the neural network. By setting $S_k = b(k)$, where S_k is the state of the k^{th} neuron and when re and im are the real and imaginary parts of the external inputs coming to the neuron, the state of the k^{th} neuron can be presented as:

$$\begin{aligned}
 S_k = f_a(& -re(r(k)h_1^c)S_{k-2} + im(r(k)h_1^c)S_{k-1}S_{k-2} \\
 & + im(r(k+1)h_1^c)S_{k+1}S_{k-1} - re(r(k+1)h_2^c)S_{k-2} \\
 & + im(r(k+1)h_2^c)S_{k-1}S_{k-2} - re(r(k+2)h_1^c)S_{k+2} \\
 & + im(r(k+2)h_1^c)S_{k+2}S_{k+1} + im(r(k+2)h_2^c)S_{k+1}S_{k-1} \\
 & - re(r(k+3)h_2^c)S_{k+2} + im(r(k+3)h_2^c)S_{k+2}S_{k+1} - re(h_1^c h_2)S_{k-3} \\
 & - re(h_1^c h_2)S_{k-2}S_{k-1}S_{k-3} + im(h_1^c h_2)S_{k-2}S_{k-3} \\
 & - im(h_1^c h_2)S_{k-1}S_{k-3} - re(h_1^c h_2)S_{k+1}S_{k-1}S_{k-2} \\
 & - im(h_1^c h_2)S_{k+1}S_{k-2} - re(h_1^c h_2)S_{k+2}S_{k+1}S_{k-1} \\
 & + im(h_1^c h_2)S_{k+2}S_{k-1} - re(h_1^c h_2)S_{k+3}S_{k+2}S_{k+1} \\
 & + im(h_1^c h_2)S_{k+3}S_{k+1} - im(h_1^c h_2)S_{k+3}S_{k+2} - re(h_1^c h_2)S_{k+3} \quad),
 \end{aligned} \tag{6}$$

where f_a is an activation function, typically a sigmoidal function or a hard limiter. In Formula (5) a signum function $f_a(x) = \text{sign}(x)$ is used.

Figure 2 illustrates a neuron according to Formula (6). The state S_k of the neuron appears at the neuron output. According to Formula (6), real and imaginary parts of the products of the received corrupted code words $r(t)$ and channel coefficients h_1 and h_2 (or their conjugates) as well as real and imaginary parts of the products of channel coefficients h_1 and h_2 (or their conjugates) are applied to the neuron as an input 200. Since the code rate of the code used was $1/2$, one bit to be encoded results in two bits when encoded. Correspondingly, in decoding one decoded bit represents two bits of the code word. Furthermore, feedback outputs of other neural network neurons, i.e. S_{k-3} , S_{k-2} , S_{k-1} , S_{k+1} , S_{k+2} , and S_{k+3} , are applied to the neuron as an input 202. The output of each neuron thus comprises one decoded bit.

In a preferred embodiment of the invention, the neuron is not fed back to itself, i.e. the value S_k is not applied to the neuron as an input. The in-

puts of the neuron are multiplied according to Formula (6). This is illustrated by dots at the intersection of the lines. White dots denote the multiplication of the inputs and black dots denote the multiplication of the inputs and the change of the sign of the product. The multiplied signals are summed 204 and applied to an activation function 206 whose output S_k 208 is the output value of the neuron.

The activation function employed can be either a hard-decision or a soft-decision function or a step function. If the result is a hard decision, the neuron outputs can have values -1 or 1. If the result is a soft decision, the neuron output can be a floating point number. Correspondingly, if the demodulator 124 produces soft bit decisions, neuron input 200 comprises floating point numbers, and if the demodulator produces hard bit decisions, the neuron input 200 comprises bits. If both of the neuron inputs comprise bits, the multiplication between the inputs can be implemented in a simple manner by digital technology.

Since the derived neural network is a so-called optimizing neural network, it may be caught in a local minimum. To avoid this, in one preferred embodiment of the invention, noise (AWGN) whose variance reduces during the stabilization stage is added to the incoming signal in the neurons. In that case, Formula (6) is of the form

$$\begin{aligned}
 S_k = f_u(& -re(r(k)h_1^c)S_{k-2} + im(r(k)h_1^c)S_{k-1}S_{k-2} \\
 & + im(r(k+1)h_1^c)S_{k+1}S_{k-1} - re(r(k+1)h_2^c)S_{k-2} \\
 & + im(r(k+1)h_2^c)S_{k-1}S_{k-2} - re(r(k+2)h_1^c)S_{k+2} \\
 & + im(r(k+2)h_1^c)S_{k+2}S_{k+1} + im(r(k+2)h_2^c)S_{k+1}S_{k-1} \\
 & - re(r(k+3)h_2^c)S_{k+2} + im(r(k+3)h_2^c)S_{k+2}S_{k+1} - re(h_1^c h_2)S_{k-3} \\
 & - re(h_1^c h_2)S_{k-2}S_{k-1}S_{k-3} + im(h_1^c h_2)S_{k-2}S_{k-3} \\
 & - im(h_1^c h_2)S_{k-1}S_{k-3} - re(h_1^c h_2)S_{k+1}S_{k-1}S_{k-2} \\
 & - im(h_1^c h_2)S_{k+1}S_{k-2} - re(h_1^c h_2)S_{k+2}S_{k+1}S_{k-1} \\
 & + im(h_1^c h_2)S_{k+2}S_{k-1} - re(h_1^c h_2)S_{k+3}S_{k+2}S_{k+1} \\
 & + im(h_1^c h_2)S_{k+3}S_{k+1} - im(h_1^c h_2)S_{k+3}S_{k+2} - re(h_1^c h_2)S_{k+3} + AWGN).
 \end{aligned} \tag{7}$$

In decoding, the following steps are generally carried out for each received code word:

1. Setting S_{-1} , S_{-2} and S_{-3} according to previous decisions. The products of the real and imaginary parts of the products of the received code words

$r(t), \dots, r(t+Tdt)$ and the channel coefficients are applied to the neuron inputs 200.

2. Initializing $S(i)$, $i = 0, \dots, T$.
3. Updating the neuron k for each $k = 0, \dots, T$ according to Formula (7).
- 5 4. Reducing the variance of the added noise. If the necessary number of iteration rounds have been carried out, the procedure is terminated, otherwise the procedure returns to step 3.

After this, the bit to be decoded is at the output of neuron 0.

Figure 3 illustrates couplings of the neural network designed for decoding code rate $\frac{1}{2}$ and constraint length 3. The demodulated code words 322 and the channel data 324 received from the channel estimation means are applied to a single neuron 306 as an input. The code words and the channel data are multiplied according to Formula (6) in a multiplier 326. Further, outputs 328, 330 from the adjacent neurons 300 to 304, 308 to 320 are applied to the neuron 306 as an input. The neuron output is the value of the activation function 326, the value being a decoded symbol. The neural network is initialized such that the previously made bit decisions become the states of the neurons -3, -2 and -1, i.e. 300 to 304. As the states of the neurons $0, \dots, T$, i.e. 306 to 320, +1 or -1 are set randomly. Thereafter, the network is allowed to stabilize. The decoded bit is the output of the neuron 306, i.e. neuron 0, or if more than one bit is decoded at a time, the outputs of the neurons $0, \dots, N$.

Duration of the network stabilization can be determined either by allowing the iterations to continue until the bit at the output does not vary or by performing a predetermined number of iteration rounds.

25 In order to avoid being caught in a local minimum, other methods together with or instead of noise addition can also be used. The number of iteration rounds to be performed can be increased. Another option is to use several parallel neural networks. When each neural network is stabilized, the best bit decision can be selected from the outputs of the networks. The selection can be made by a majority decision, for instance, selecting the bit decision that occurs most frequently at the outputs of the parallel neural networks. It is also possible to re-encode the decoding result and compare the obtained result with the received code word, i.e. to select a result that minimizes the function

$$35 \quad \frac{1}{T_b} \sum_{s=0}^{T_b} \|r(s) - r_{re}(s)\|^2. \quad (8)$$

Decoding can also be accelerated such that several bit decisions of the adjacent neurons, i.e. neurons S_0, \dots, S_N , are extracted simultaneously from the neural network. When decoding the subsequent bits, the states of the neurons S_{-L+1}, \dots, S_{-1} have to be initialized to have the preceding values.

5 The method of the invention can also be applied when the code to be decoded is a punctured code.

The solution of the invention can also be advantageously applied such that fast channel estimates are utilized. Each symbol can be expressed using symbol-specific channel data, even though the convolutional code extends over a plurality of symbols.

Let us study next the complexity of the solution of the invention mainly from the viewpoint of implementation. When the constraint length L increases, more neurons are needed in the neural network and the duration of stabilization increases, and consequently the number of multiplications increases. It is known that the complexity of a Viterbi decoder increases exponentially as the constraint length L increases. However, the complexity of the decoder implemented by means of the neural network of the invention increases much more slowly. When the constraint length is L and the number of channel taps is H , each neuron of the neural network is coupled to $L+H-2$ preceding and $L+H-2$ subsequent neurons and to L inputs. If it is assumed that the decoding delay is $5L$, as it generally is, the decoder requires $5L$ neurons and thus the maximum number of connections is in the order of $O(H^2L^2n)$, where n is the number of generators.

The number of multiplications to be performed in one neuron depends on the code generator used, but it is $\leq O(nH^3 + nH^2L + nHL^2)$. Thus, during one iteration round, when each neuron is updated once, the number of multiplications to be performed is $O(nLH^3 + nH^2L^2 + nHL^3)$. This shows how the complexity in the solution of the invention increases considerably more slowly than in connection with the Viterbi algorithm as L increases.

30 Figure 4 illustrates the implementation of the neuron of the neural network according to the invention. The neuron state at the neuron output is S_k . Interfered code words to be decoded, typically composed of bits, and channel data are applied to the neuron as an input 400. The number of bits in each code word depends on the code rate of the code used. Further, feedback outputs of other neural network neurons are applied to the neuron as an input 35 402. In a preferred embodiment of the invention, the neuron does not com-

prise feedback to itself, i.e. the value S_k is not applied to the neuron as an input, but feedback to itself is possible if necessary.

The inputs of the neuron are multiplied in a matrix 404. The matrix couplings are defined by the convolutional code and channel model used, and
5 the formula corresponding to the couplings can be derived as illustrated in the above examples. The inputs can be multiplied with or without a change of sign. The multiplied signals are applied to a summer 406 where they are summed and the summed signal is further applied to a calculation means 408 which calculates the activation function and whose output S_k 410 is the output
10 value of the neuron. The calculation means 408 can be implemented by a gate array or a processor and software.

The function producing either a hard or a soft decision can thus be used as the activation function. If the result is a hard decision, the neuron outputs can have values -1 or 1. If the result is a soft decision, the neuron output
15 can be a floating point number. Correspondingly, if the demodulator of the receiver produces soft bit decisions, floating point numbers are at the neuron input 400, and if the demodulator produces hard bit decisions, bits are at the neuron input 400. If both of the neuron inputs comprise bits, the multiplication between the inputs can be implemented in a simple manner by digital technology.
20

In one preferred embodiment of the invention, the neuron comprises a noise generator 412 and the generated noise is applied to a summer 406 where it is added to a sum signal. Variance of the noise to be generated reduces during the stabilization of the neural network, as described above.
25 The noise can be Gaussian noise or binary noise. The binary noise is easier to implement, but on the other hand, a suitable number of Gaussian noise samples can be stored in a memory element and this noise can be reused. Noise generation can be implemented in manners known to the person skilled in the art.

30 The structure of the decoder of the invention is studied next from the viewpoint of implementation. Figure 5a illustrates the structure of the neural network of the invention. The neural network comprises a set of neurons 500 to 510, the implementation of which is described above. Code words 512 that are applied to a shift register 514 are applied to the neural network as an
35 input. The latest $T + L$ code words, where T is the decoding delay, are stored in the shift register. Upon completing the decoding of a code word, the de-

coded code word is deleted from the register, the code words in the register are shifted one step forward and a new code word is adopted in the last storage location 516 of the register. Operationally, the shift register is connected to the inputs of the neural network neurons 500 to 510. Channel data 517 from the channel estimation means is further applied to the neural network as an input. The channel data is applied to a buffer memory 518. The neural network further comprises a buffer memory 520 whose input comprises the outputs of the neurons 500 to 510 and whose buffer memory output is operationally connected to the inputs 724 of the neural network neurons. The number of the neurons depends on the characteristics of the convolutional code to be decoded.

The outputs of the neural network neurons are thus operationally connected to the buffer memory 520, from which the outputs are fed back 522 to the inputs 524 of the neurons. The neuron inputs comprise switching means 524 by which only the necessary code words and feedbacks are applied to each neuron, in accordance with a formula (as Formula (7) above) derived on the basis of the convolutional code to be decoded and the channel data. The switching means 524 can be implemented by means of a switching matrix or a discrete logic circuit. The bit to be decoded is obtained from the output of the calculation means 526 calculating the activation function. The calculation means 526 can be implemented by separate logic circuits or a processor and software.

The neural network decoder comprises a control unit 528, to which is applied a clock signal 530 as an input from the timing information regenerator of the receiver, as is described in connection with Figure 1. The control unit takes care of the decoder timing, the control of the network stabilization time and controls the operation of decoding.

Figure 5b illustrates a second preferred embodiment, in which several bit decisions of the adjacent neurons, i.e. neurons $k \dots k+n$, are extracted simultaneously from the neural network. In that case, the equipment comprises a plurality of calculation means 526a, 526b. Also in that case, upon completing the decoding of the code words, n decoded code words are deleted from the register, the code words in the shift register 514 are shifted a corresponding number n of steps forward and n new code words are adopted in the register.

Figure 6 illustrates yet another embodiment of the invention, in which decoding is performed by means of a plurality of parallel neural networks. In that case the code words and the channel data 600 are applied to two or more neural networks 602, 604 arranged in parallel. The outputs of the calculation means 606, 608 of the neural networks are applied to a decision unit 610 where the best estimate for a decoded bit is selected from the outputs of the parallel neural networks. The decision unit can be advantageously implemented by means of separate logic circuits or a processor and software. The decision on the best estimate can be implemented in a variety of ways. For instance, from the outputs of the parallel neural networks, it is possible to select the value that occurs at most outputs to be a decoded bit. The decoded bit can also be re-encoded and from the outputs of the parallel neural networks the value that is found correct by re-encoding is selected to be the decoded bit. For the sake of clarity, re-encoding is not shown in the figure but it can be implemented in manners known to the person skilled in the art.

The decoder of the invention can be readily constructed by semiconductor technology, since the neurons of the neural network are very similar to one another in structure, only the input couplings vary. Consequently, to design and implement even a large network does not involve a great amount of work. Testing of the completed circuit is also simple to implement. The solution can also be advantageously implemented as a VLSI circuit, which makes it fast. It should also be noted that even though the structure of the decoder of the invention is simple and easy to implement, its performance is close to the theoretical maximum.

Even though the invention is described above with reference to the example of the accompanying drawings, it is obvious that the invention is not restricted thereto, but it can be modified in variety of ways within the scope of the inventive idea disclosed in the attached claims.

CLAIMS

1. An arrangement for decoding and channel correcting a convolutionally encoded signal received over a transmission path, the signal comprising code words and the arrangement comprising a neural network which comprises a set of neurons which comprise a set of inputs and an output, received code words (400) and at least some of the output signals of the neural network neurons being connected to the inputs of said neurons, the neurons comprising means (404) for multiplying at least some of the neuron inputs prior to combining means (406) and the arrangement comprising means (123) for estimating the transmission channel, **characterized** in that estimated channel data (517) is further connected to the inputs of the neurons and that a predetermined neuron is arranged to give an estimate of a channel-corrected and decoded symbol in its output signal.

2. An arrangement as claimed in claim 1, **characterized** in that the neuron comprises means (404) for multiplying at least some of the neuron inputs and for changing the sign of the product prior to the combining means (406).

3. An arrangement as claimed in claim 2, **characterized** by comprising a shift register (514) having the received code words as an input, and a memory element (518) having the estimated channel data (517) as an input, said shift register and memory element being operationally connected to the inputs of the neural network neurons, and a buffer memory (520) having the outputs of the neurons as an input, the output of said buffer memory being operationally connected to the inputs of the neural network neurons.

4. An arrangement as claimed in claim 3, **characterized** by comprising means (128, 528) for conducting the received code word set (512) to the shift register (514), the estimated channel data (517) to the memory element (518), for setting initial values of the neurons and for conducting the estimate of the decoded symbol to the output of the network after the network has reached a stabilized state, for updating the code word set and the estimated channel data in the shift register and for repeating the above-described operation until the desired number of code words are decoded.

5. An arrangement as claimed in claim 2, **characterized** in that the neuron comprises means (412) for adding noise to the result of combination in the combining means.

6. An arrangement as claimed in claim 1, **characterized** in that the output signal of a plurality of predetermined neurons comprises an estimate of the corresponding number of decoded and channel-corrected symbols.

- 5 7. A method for decoding a convolutionally encoded signal received over a transmission path, the signal comprising code words and the method comprising the steps of
- estimating a transmission channel,
- performing the decoding by means of an artificial neural network,
- 10 the neural network comprising a set of neurons which comprise a set of inputs and an output,
- conducting the received code word set (512) to the inputs of the neuron,
- multiplying at least some of the inputs of the neuron by one another,
- 15 combining some of the inputs of the neuron after multiplication,
- feeding some of the output signals (520) of the neural network neurons back to the inputs of each neuron,
- setting the initial values of the neurons and
- allowing the network to stabilize,
- 20 **characterized** in that further in the method
- the multiplication of the neuron inputs depends on the convolutional code used in signal encoding and the estimated channel and that
- the estimate of the decoded and channel-corrected symbol is conducted from the output signal of a predetermined neuron to the output of the
- 25 network after the network has reached the stabilized state,
- the code word set in the shift register is updated
- and the above-described four last steps are repeated until the desired number of code words are decoded.

- 30 8. A method as claimed in claim 7, **characterized** in that at least some of the neuron inputs are multiplied in the neuron and the sign of the product is changed prior to the combination.

9. A method as claimed in claim 7, **characterized** in that, in connection with combining the neuron inputs, noise is added to the outcome of the combination.

10. A method as claimed in claim 9, **characterized** in that noise is added such that the variance of the noise approaches zero as the network approaches stabilization.

11. A method as claimed in claim 7, **characterized** in that
5 the network is allowed to stabilize for a predetermined period of time.

12. A method as claimed in claim 7, **characterized** in that the network is allowed to stabilize until the values of the output signals of the neurons do not vary.

13. A method as claimed in claim 7 or 9, **characterized** in
10 that after combining the neuron inputs the signal is applied to an activation function, at the output of which the output signal of the neuron is obtained.

14. A method as claimed in claim 13, **characterized** in that from the combined signal the activation function makes a hard decision.

15. A method as claimed in claim 13, **characterized** in that
15 from the combined signal the activation function makes a soft decision.

16. A method as claimed in claim 13, **characterized** in that the activation function is a step function.

17. A method as claimed in claim 7, **characterized** in that the received code word set (512) is conducted to the neuron inputs by means
20 of a shift register (514).

18. A method as claimed in claim 17, **characterized** in that the estimate of the decoded symbol is conducted from the output signal of a plurality of predetermined neurons to the output of the network after the network has reached the stabilized state and that the code word set in the shift
25 register is updated with the same predetermined number of code words.

19. A method as claimed in claim 7, **characterized** in that the channel estimate is updated such that an estimate of each symbol is calculated by using a dedicated channel estimate.

20. A method as claimed in claim 7, **characterized** in that
30 decoding is performed by means of a plurality of parallel neural networks and that from the outputs of the parallel neural networks the best estimate is selected to be a decoded bit.

21. A method as claimed in claim 20, **characterized** in that from the outputs of the parallel neural networks the value that occurs in most
35 outputs is selected to be the decoded bit.

22. A method as claimed in claim 20, **characterized** in that from the outputs of the parallel neural networks the value that is found correct by re-encoding is selected to be the decoded bit.

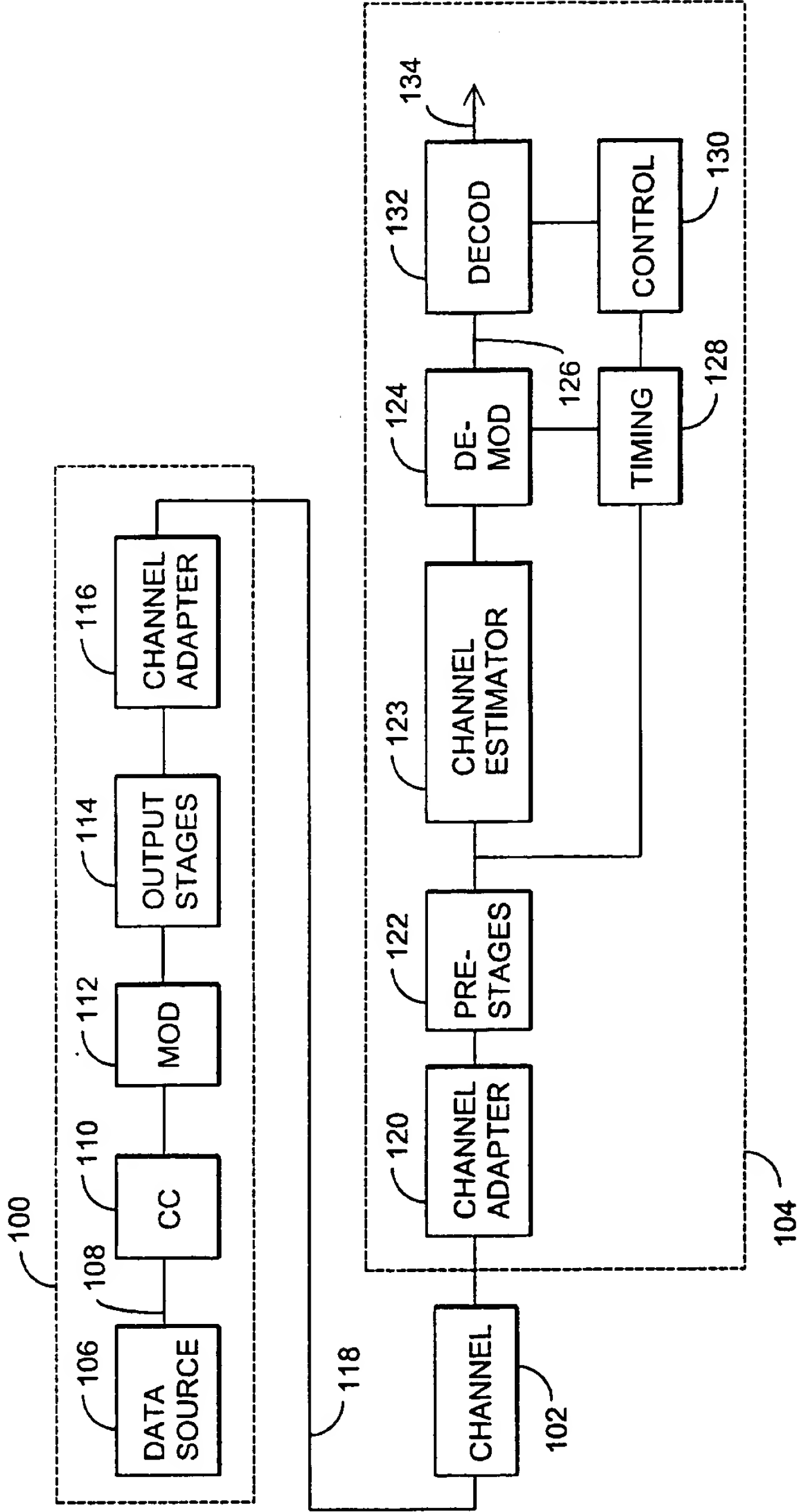


Fig. 1

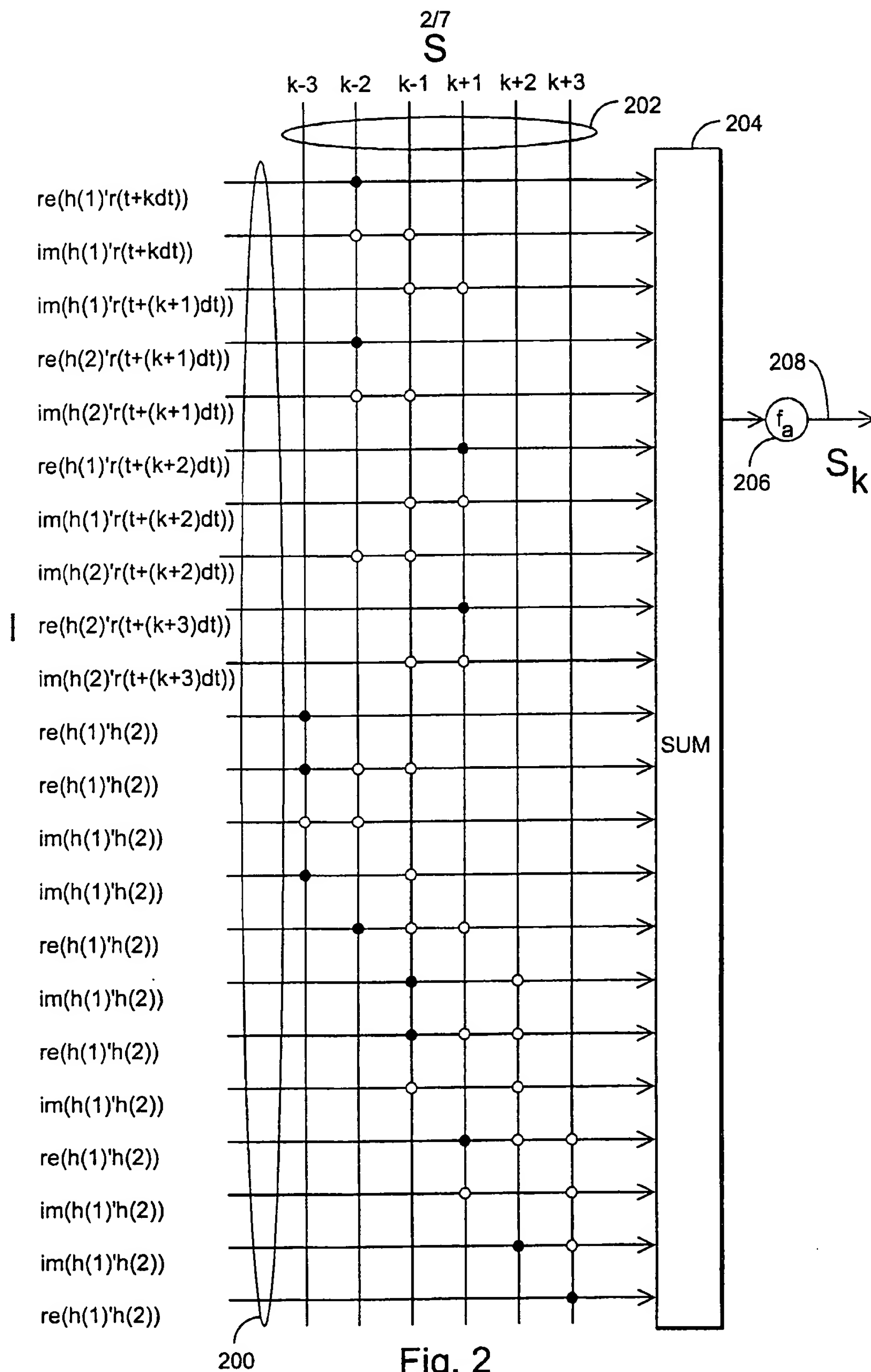


Fig. 2

3/7

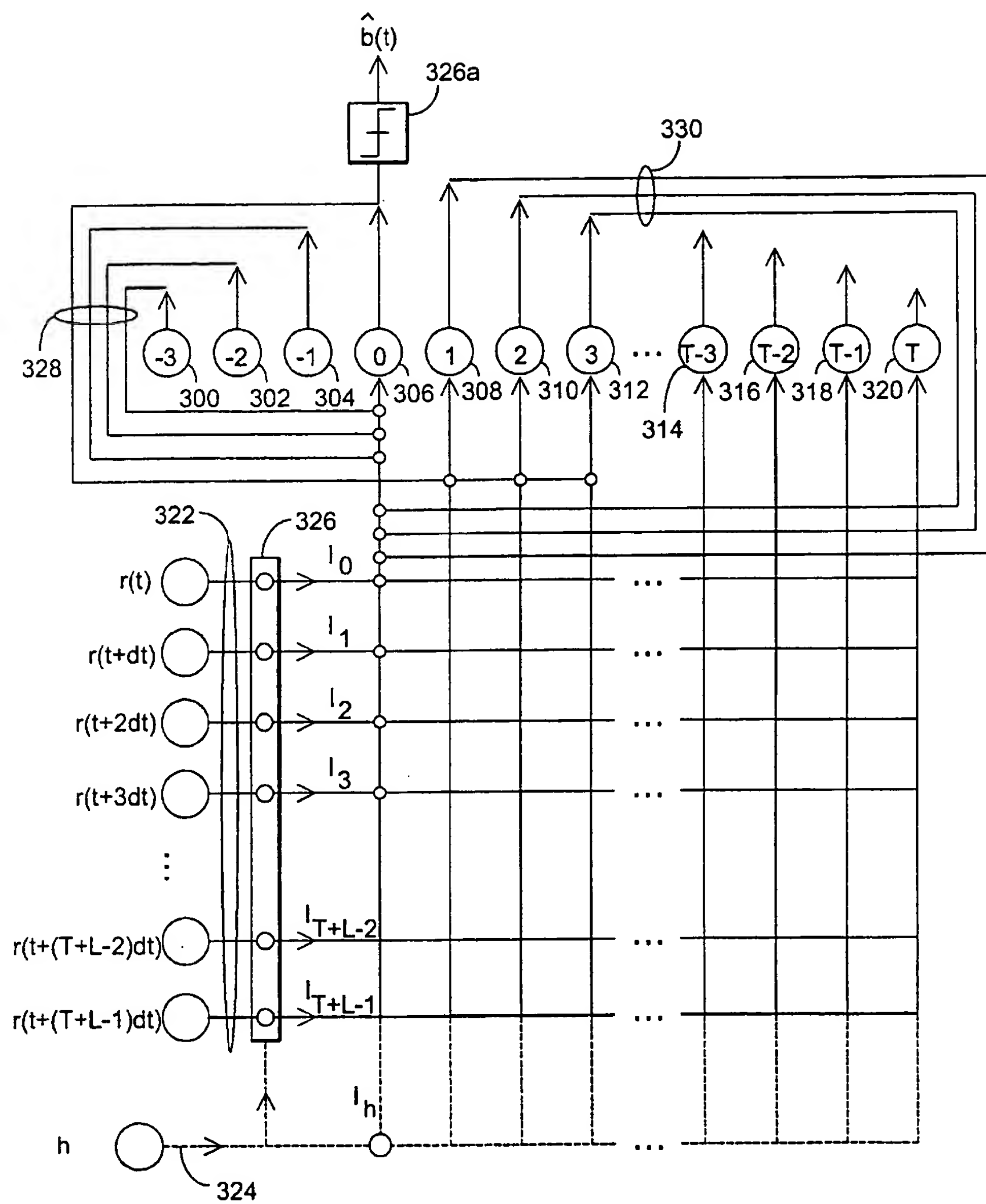


Fig. 3

4/7

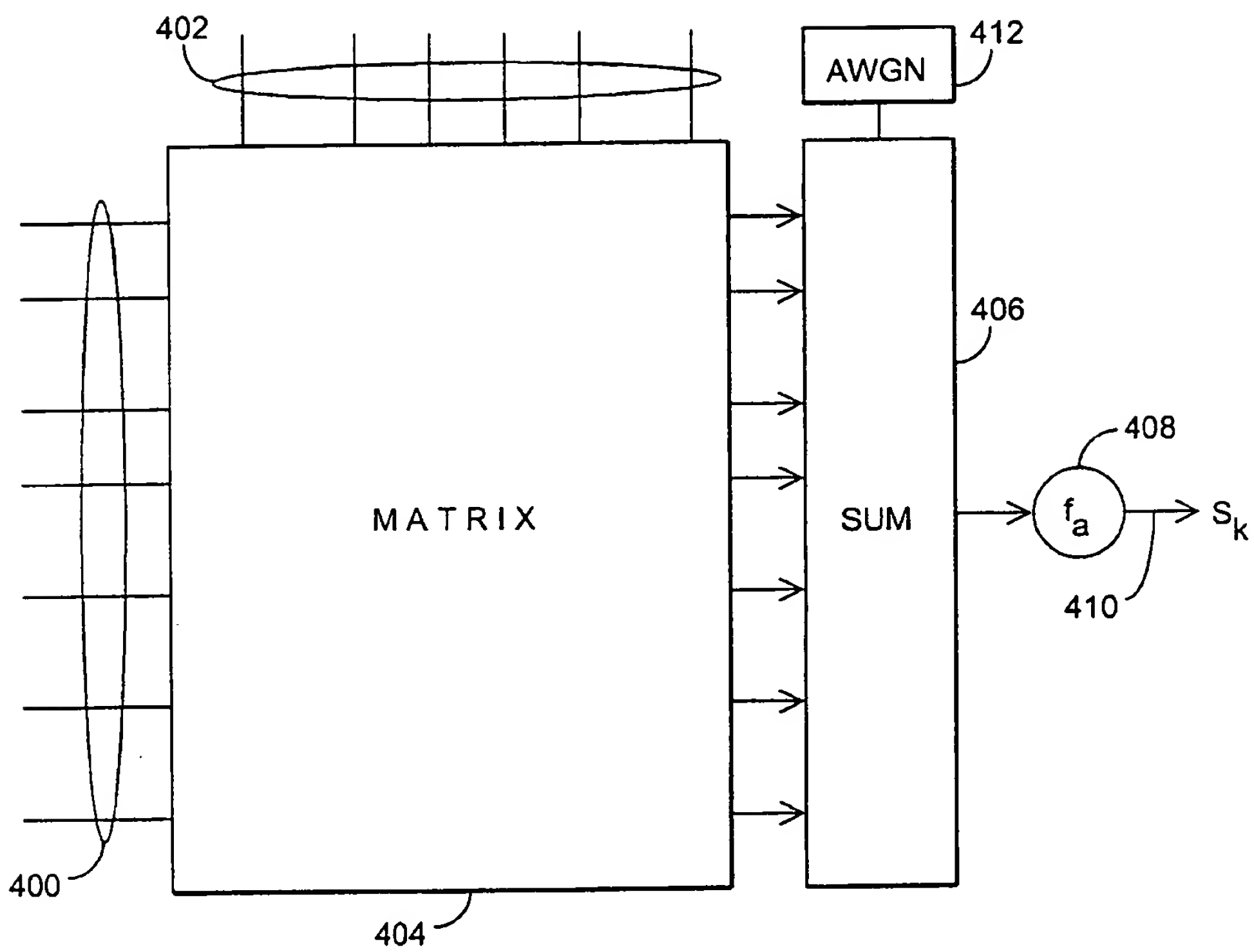


Fig. 4

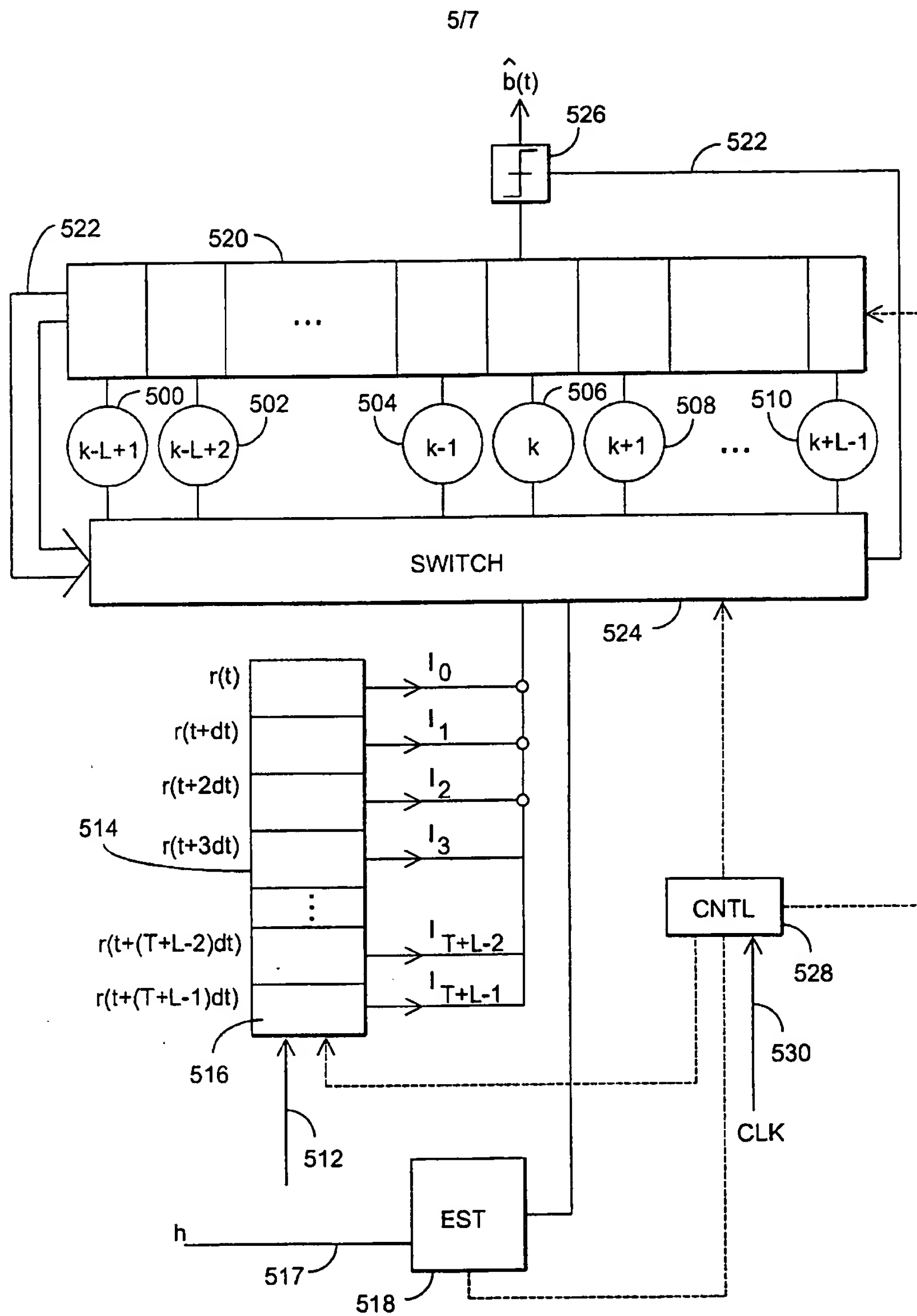


Fig. 5a

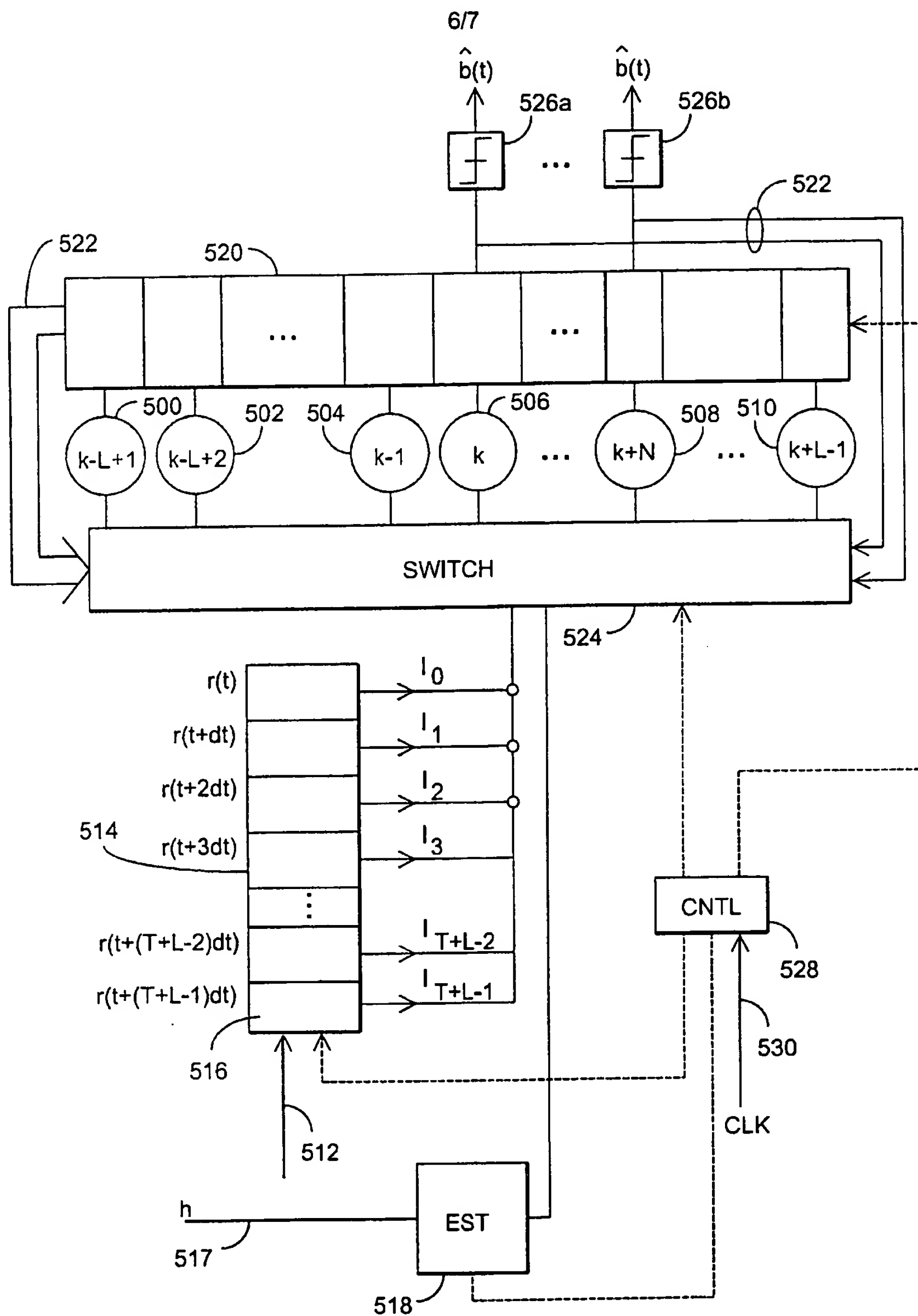


Fig. 5b

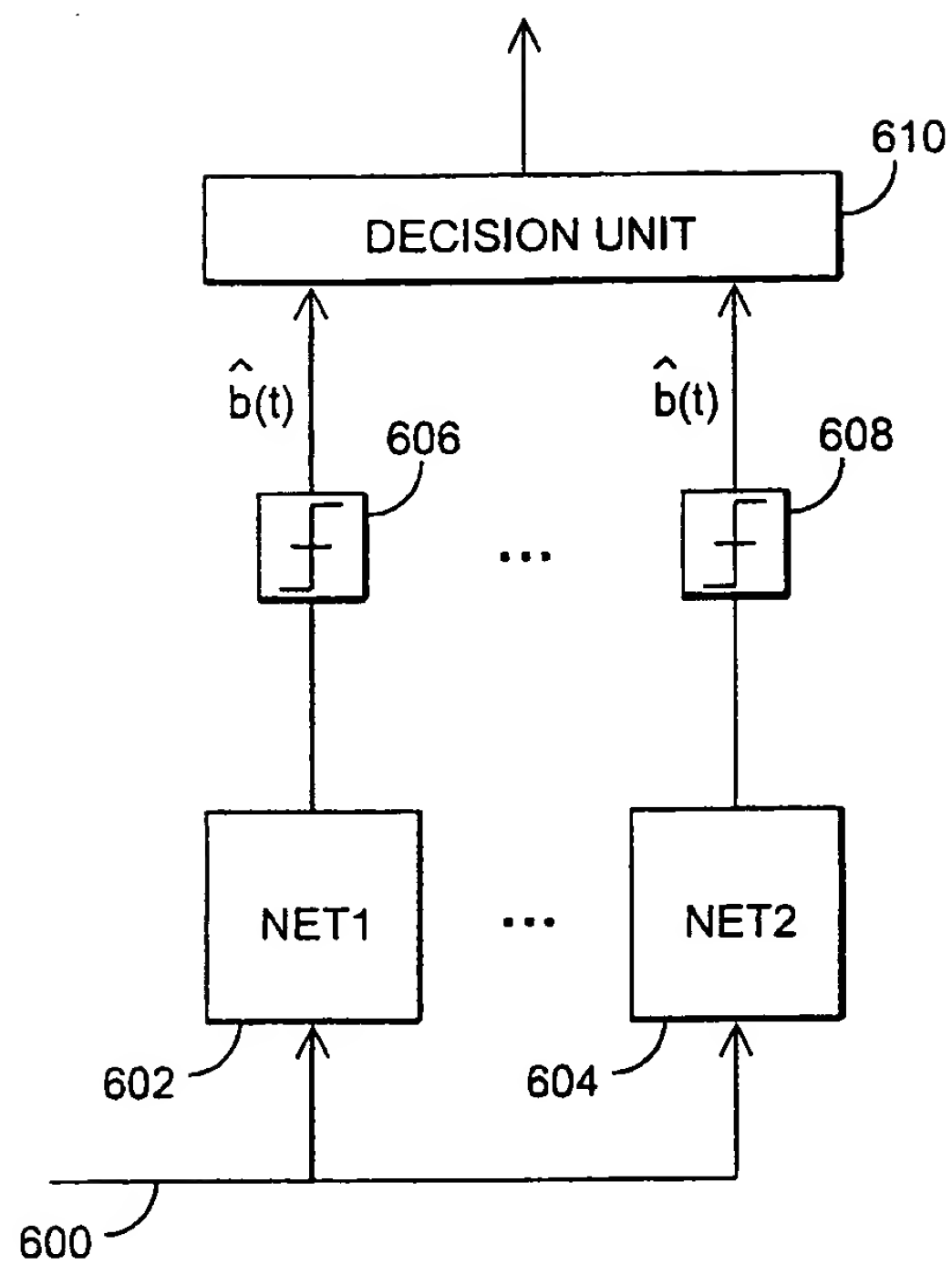


Fig. 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI 00/00457

A. CLASSIFICATION OF SUBJECT MATTER

IPC7: H03M 13/23, H03M 13/37

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC7: H03M, H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5764858 A (B.J. SHEU ET AL.), 9 June 1998 (09.06.98), column 3, line 10 - column 4, line 2; column 1, line 17 - line 35; column 8, line 30 - column 10, line 12, figure 3a --	1-22
X	IEEE Transactions on Neural Networks, volume 56, Nov. 1994, Pages 982-988, Khalid A. Al-Mashouq et al: "The use of neural nets to combine equalization with decoding for severe intersymbol interference channels", see section V.B; abstract --	1-22

☒ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

29 Sept 2000

Date of mailing of the international search report

04 -10- 2000

Name and mailing address of the ISA/

Swedish Patent Office

Box 5055, S-102 42 STOCKHOLM

Facsimile No. +46 8 666 02 86

Authorized officer

Peder Gjervaldsaeter/AE

Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI 00/00457

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	International Conference on Neural Networks, 1997, Volume 4, 1997, Pages 2080-2084, Kimmo Raivio et al: "Performance of two neural receiver structures in the presence of co-channel interference", see section 1;3 --	1-22
A	1997 13th International Conference on Digital Signal Processing Proceedings, 1997. DSP 97, Volume 1, 1997, Pages 287-290, Lorenzo Favalli et al: "Non-linear mobile-radio channel estimation using neural networks", see abstract; section II --	1-22
A	IEEE International Conference on Neural Networks, 1995. Proceedings, volume 1, Pages 449-454, Werner G. Teich et al: "A novel decoder structure for convolutional codes based on a multilayer perceptron", see section 3 --	1-22
P,A	EP 0967731 A2 (NOKIA MOBILE PHONES LTD.), 29 December 1999 (29.12.99), see the whole document -- -----	1-22

INTERNATIONAL SEARCH REPORT
Information on patent family members

01/08/00

International application No.

PCT/FI 00/00457

Patent document cited in search report			Publication date	Patent family member(s)	Publication date
US	5764858	A	09/06/98	NONE	
EP	0967731	A2	29/12/99	FI 981168 A	27/11/99